TITLE: A response to the Forensic Genetics Policy Initiative's Report "Establishing Best Practice for Forensic DNA Databases"

Dear Editors,

Over the last few decades, a series of concerns have been raised about DNA profiling and databasing in the criminal justice system, in particular with relation to privacy and discrimination. There have also been calls to ensure responsible governance, ongoing consideration of safeguarding human rights, and appropriate multi-disciplinary and public debate relevant to ethical, social and political issues. The Forensic Genetics Policy Initiative's 2017 Report: "Establishing best practice for forensic DNA databases" [1] provides an important response to these concerns.

As a group of experts from a number of disciplines and backgrounds (including genetics, political science, sociology, law and ethics) and incorporating members of a range of organisations including the Public and Professional Policy Committee of the European Society of Human Genetics and the PHG Foundation, we commend the ambition of this 2017 Report. We also particularly applaud the Report's breadth and depth of focus on the entire chain of responsibility from crime scene examination to the court system, including wider infrastructural elements. The Report is absolutely right to highlight that forensic databases should focus on DNA of convicted persons; that forensic use of DNA without the subject's full, informed consent can only be justified in limited circumstances, such as to solve a very serious crime; and that non-forensic DNA databases should, in general, not be accessible for forensic purposes. We also underscore the issues raised in the Report about the risk of over-reliance on DNA evidence - perhaps, as we would suggest, because we live in a society where technological evidence is often trusted more than human judgement; and agree with the Report's emphasis on the continued need to balance expected benefits with the risk of misuse of information held in forensic databases.

At the same time we believe some issues within the Report deserve more specific and more systematic attention. We provide a brief overview of these here, seeking to contribute to setting the agenda for what we hope will be broad discussions on the governance of forensic genetics in the criminal justice system involving multiple disciplines and stakeholders.

Blurring boundaries: DNA-based information inside and outside of forensic databases

Many countries currently have functioning national forensic DNA databases which store the DNA profiles of suspected and/or convicted criminals in the form of short tandem repeats (STRs).[i] Some jurisdictions' databases also include DNA profiles from other individuals such as victims or volunteers. When there are no known suspects of a crime, traditional forensic DNA testing uses these databases to compare an STR profile obtained from a crime scene with all STR profiles in the database to see whether there is a match. If there is, this *could* mean that the person whose profile matches a crime scene profile is involved in some way with the crime (for example, by committing or aiding in the crime, or being present during the crime), but this would need to be established with further evidence; a profile match alone is not sufficient proof. STR profiles used in forensics are derived from markers located in non-coding DNA regions and as such, it is perceived that no information regarding disease- or personality-related characteristics can be inferred.[ii] With the advent of high-throughput next-generation sequencing (NGS) or massive parallel sequencing (MPS), the efficient generation of data at the nucleotide level beyond that of STR profiles alone has allowed laboratories to produce much more wide-ranging DNA information - for example, single nucleotide polymorphism (SNP) data from within STR markers, which can lead to greater discrimination between alleles. This has had implications for the development of new forensic genetic approaches, one example of which is forensic DNA phenotyping (FDP) [2].

[i] Short sequences of DNA situated in defined places (loci) across the genome, which are tandemly repeated numerous times. The set number of times an STR is repeated can differ between individuals, and between each individual's STRs, and it is these numbers which are stored in forensic DNA databases.

[ii] Though we note that the separation between coding and non-coding DNA, which so far has served as an "ethical boundary" to determine what DNA information could be used for forensic purposes and what could not, is becoming increasingly blurred.

FDP allows for the probabilistic inference of likely phenotypic characteristics from DNA, such as age, appearance and ancestry. Rather than being stored in national forensic DNA databases, in countries where FDP is used for specific criminal investigations, SNP-based information is stored de-centrally in laboratories performing the analysis. This leads to a scenario in which DNA-based information in criminal investigations is used in wider contexts than centralised STR-databases alone. Identifying what responsible governance should look like for forensic databases that *do not* use national centralised databases[iii] therefore now requires consideration of how such findings should be stored; who should have access to them; and how findings (which are highly probabilistic and predictive, and raise issues of discrimination) should be communicated with, and inform, operative police work.

The implications of NGS or MPS raise other new regulatory, ethical, and social questions.[iv] For example, current pushes for more public donation of genomic (and linked clinical) data to biobanks mean that non-forensic DNA databases are becoming larger, both in terms of the types of data included, and the number of people whose data are stored in them. Especially in high profile criminal cases it will be difficult to resist the political push to make data in these databases, which is curated for medical research, open to use in criminal investigations. Whilst not directly applicable to biobanks, a recent case in the United States exemplifies this issue. Here, law enforcement identified a familial match [3] of a suspected perpetrator by searching a free online genetic genealogy database, GEDMatch, where people interested in finding genetic relatives upload their DNA profiles that they have previously obtained from private DNA companies.[v] Despite the relief that this case could be solved, commentators were concerned about police accessing a database that

[iii] On a minor note, the document states that in the jurisdictions in which FDP is currently being used, the technology is being sold as a commercial service to the police. This is inaccurate - in most jurisdictions in which FDP is being used, the police use their own, or academic, forensic centres who have the expertise to perform such tests.

[iv] https://www.gov.uk/government/consultations/ethical-dimensions-of-next-generation-sequencing

[v] http://www.bbc.co.uk/news/world-us-canada-43916830

people had contributed to who did not have law enforcement use in mind. This raises deep questions about function creep [4], about meaningful informed consent,[vi] and about the requirement for public deliberation about what - and how - boundaries between functional purposes of databases containing personal information and/or public databases containing DNA should be maintained. This deliberation has become increasingly urgent following the declaration of cooperation signed by 13 EU member states to reach a shared collection of one million sequenced genomes accessible in the EU by 2022.[vii]

The context of big data and data mining

The Report regards forensic databases in isolation and not as part of a wider context of big data, automated processing and data mining. By doing so, the Report misses the present key challenge for the use of DNA information for crime prevention: the integration of DNA-based information, biometric data, and also data from sensors in public and personal domains (not only CCTV but also "smart speakers" etc). Data from Google Home and Amazon Echo have already started to be used in ongoing investigations.[viii] Different datasets, taken together, can be used to detect patterns of particularly "risky" individuals, and this information can then be used preemptively.[ix] For example,

[vi] The "Best Practice" Report states that "*Best practice for police access to stored genetic information...requires strict oversight: including not only authorisation by a court, but also carefully defined guidance on the circumstances in which such requests can be granted, and how much data can be revealed. Further, information needs to be provided to people who take part in such databases so they are aware*". Such information is currently provided to customers of commercial biobanks and DNA testing services. However, when the terms of service of biobanking and direct-to-consumer (DTC) genetic testing websites were studied, policies mentioned that genetic information may be shared with "authorities" via court order (Niemiec, E., & Howard, H. C. (2016). Ethical issues in consumer genome sequencing: Use of consumers' samples and data. *Applied & Translational Genomics*, *8*, 23–30.). 23andme's website states that "*We will not use your sensitive information without your consent unless: (i) the information has been anonymized or aggregated so that you cannot reasonably be identified as an individual; or (ii) a legal obligation requires us to use it in some way e.g. a court order requires us to disclose the information*, but it is questionable how many people read this information, which is often in the small print of the terms of service (https://www.23andme.com/en-gb/about/privacy/).

[vii] http://euapm.eu/pdf/EAPM_Declaration_Genome.pdf

[viii] http://www.consumerwatchdog.org/sites/default/files/2017-12/Digital%20Assistants%20and%20Privacy.pdf

[ix] https://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/

using combinations of datasources in this way could impact on eligibility for other secondary uses, such as insurance. Whilst such scenarios may be some way off, they are already beginning to be considered in policy circles.[x] We stress that these considerations include discussions relating to data protection, data contextualisation, education, and the safeguarding of civil liberties.

Exchanging forensic DNA information across borders

We applaud the Report for referring to the Prüm regime, which allows the exchange of forensic DNA information across the national databases of 24 EU Member States, and which will remain of great relevance for Europe in the foreseeable future. The Report, however, misses a key aspect of the Prüm regime, namely that it has led to *less* personal data crossing borders, rather than more. This might be counterintuitive at first sight. But prior to Prüm, countries sent entire files including personal data from one country to another when they collaborated on an investigation. Now with Prüm, no personal information is made available unless the information held in both countries' databases indicates a 'match'. Only then will personal information cross borders. This underscores the importance of well-designed (in terms of privacy by default and privacy by design) systems for digital data exchange which can *reduce* privacy risks if implemented well (and if there are no other practices that increase risks to privacy).

Data protection

We are concerned with the Report's confusion regarding relevant data protection frameworks for DNA profiling and databasing: neither the Data Protection Directive,[xi] nor its successor, the General Data Protection Regulation (GDPR), are applicable to data processing by competent authorities for law enforcement purposes. Rather, a *lex specialis* is - which will be replaced by the

[x] The renaming of the UK DNA Database Ethics Group into "Biometrics and Forensics Ethics Group" reflects this: https://www.gov.uk/government/organisations/national-dna-database-ethics-group

[xi] Directive95/46/EC on the protection of individuals with regard to the processing of personal data (PII (US)) and on the free movement of such data

EU Police Directive from May 2018.[xii] This means that some of the notice and consent requirements that the GDPR prescribes (and that the Report considers essential) will not be granted, as the Police Directive provides exceptions in cases where notification would impede public interest in an ongoing investigation.

The commercialisation of forensic science

Whilst the Report mentions the overarching commercial interests in forensic DNA technology development and use, we believe that this requires more scrutiny. In particular, we argue that technology development for law enforcement purposes needs to be something that is a public function and should not rely exclusively on commercial providers. Accountable public bodies need to play an important role, not only in setting standards for technology validation and deployment but they also need to provide services themselves where there is strong public interest that service provision is committed solely to public interest and not commercial profits.

Points to consider

Moving forward, in line with/building on the initiative of the report on *Establishing Best Practice for Forensic DNA Databases*, and given the issues discussed in this response, we urge that any discussion of human rights safeguards in the context of forensic genetics is opened up to the debate about the governance of forensic genetics and its associated social, ethical and regulatory concerns in this wider sense. In particular we suggest the following points to consider:

- *Next Generation Sequencing*. It will be increasingly hard to insist that only 'non-coding' DNA can be used for criminal investigation. We thus welcome ongoing initiatives to

[xii] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA

systematically consider the ethical and social dimensions of these practices and contribute to better policies to regulate the use of DNA information.

- *Growing possibilities of integrating data from several sources*: Instead of treating forensic DNA information as a system on its own, consider the collection and use of data in forensic DNA databases as part of a larger data ecosystem with greater possibilities to integrate different types of data.
- *Use of distributed data*: Be attentive to the ethical challenges related to contexts where DNA and other data are used to prevent, solve, and punish crimes do not come from centralised databases but exist decentrally in local laboratories, people's personal devices, and commercial companies.
- *New accountability deficits*: If wider sources of data and information are used than are stored in centralised DNA databases, this means that control over what data is included, over quality control etc. may not lie in the hands of public authorities but private corporations or even single individuals. This can create accountability and transparency deficits that need to be attended to.
- *Commercial interests in DNA information for preventing and investigating*: The changing landscape of actors in the development, provision, and monitoring of tools and services related to forensic DNA merits a fresh assessment of the effects of commercial interests in this field, and how they can be made transparent and accountable to the public.
- *Stimulate regulation, awareness and debate on potential uses of DNA-based information obtained for non-forensic purposes*, such as in the context of health care, research or ancestry testing that may increasingly become available in the public domain.
- *Ongoing professional guidance:* As the applications of forensic genetic technologies expand and diversify, education about when to use the technologies, and their potential benefits, limitations and uncertainties is essential for all those working in the criminal justice system.

References

[1] *Establishing best practice for forensic DNA databases* (2017). Report by the Forensic Genetics Policy Initiative http://dnapolicyinitiative.org/wp-content/uploads/2017/08/BestPractice-Report-plus-cover-final.pdf

[2] Kayser, M. (2015). Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. *Forensic Sci Int Genet*. Sep;18:33-48.

[3] Maguire C. N., McCallum L. A, Storey C., Whitaker J. P. (2014). Familial searching: a specialist forensic DNA profiling service utilising the National DNA database to identify unknown offenders via their relatives--the UK experience. *Forensic Sci Int Genet*. Jan;8(1):1-9.

[4] Dahl, J.Y. and Sætnan, A.R. (2009). "It all happened so slowly"– On controlling function creep in forensic DNA databases. *International journal of law, crime and justice*. 37(3): 83-103.